# COMPUTER SIMULATION OF SMALL GROUP DECISIONS

A. P. Hare

R. Richardson

H. Scheiblechner

Research Memorandum no. 13

February 1968

The research team wants to express its thanks to the following persons contributing to the solution of technical and scholarly problems in connection with this research problem:

Prof. Dr.E.F. Winter (director) and
Dr. techn. J. Bomze (vice-director),
Dr. G. Wieser (section chief, sociology),
Dr. R. Stöckelle (computer center),
the secretarial staff of the institute and
Mr. J. Auner

The authors:

A. P. Hare, visiting professor of the Institute for Advanced Studies
Haverford College
Haverford, PA 19041

R. Richardson
Northeastern University
Boston, Mass.

H. Scheiblechner
Institute for Advanced Studies
Stumpergasse 56
Vienna

A comprehensive scheme for the simulation of interpersonal behavior in small groups has been outlined by Bales, Couch, and Stone in their description of the "Interaction Simulator" (Bales, Couch and Stone, 1961). The process of simulation would begin after a set of subjects are given a battery of personality and performance tests. Before they come together for a group discussion, the ideal computer simulation would indicate which subject will choose each of the seats around the table, to whom each will speak and in what order, and how the problem solving will proceed as the group reaches one or more decisions. This simulation would then be compared with the outcome of the actual group discussion. So far relatively little of this scheme has been programmed for the computer although Bales, Couch, Stone and others have been able to generate theories and methods which bring the task much closer to realization than when it was first proposed (See Bales, 1968, Couch, 1960, Stone et al, 1966.)

The present simulation models are concerned with only one aspect of the more general problem, namely the process by which group members pool opinions to form a group judgement. These models, which pose different methods of averaging individual judgements, are based on earlier models which used some of the same data to test the simulations (Hare, 1961, Hare and Richardson, 1966). In addition to the data from groups composed of American college students used to test the previous simulations, new data have been obtained from groups of Austrian college students solving a set of similar problems.

## The process to be simulated

The process to be simulated is the formation of a group opinion about a subject, once individual members have formed opinions about the same subject. Since several pre-computer studies of group decisions have indicated that the "pooling" of individual opinions may represent accurately the results of group discussion, it is expected that this process will involve some form of averaging such as taking the mean, median, or mode.[1)]

The group to be simulated is a five-man laboratory discussion group of American and Austrian college undergraduates. They are seated at a table with three members along one side, one at each end, and the fourth side open towards the experimenter who records the interaction rate for each member and monitors the task. The task is described in some detail in a previous article (Hare, 1961). In brief it consists of giving each of the group members a question from a questionnaire concerned with valueorientations, together with the answer given by an "unknown subject".[2] The group members are told to pool their information about the unknown subject and then predict his answers to a set of ten or more questions. After each prediction the unknown subject's actual answer is revealed and is discussed by the group before the members make their next prediction. Before each trial the group members record their individual predictions.

## Three models of group decisions

The first model for group decisions assumed that group members took an average of their individual opinions and that this average was best represented by the mean (Hare, 1961). The computer program began with the five individual predictions as inputs. The group decision was simulated by taking the mean of the five predictions. The predictions for the American groups werde made on a seven point scale where -3 represented "strongly disagree", + 3 represented "strongly agree", and 0 represented no opinion. Since group members were required to give some opinion, the computer program provided for an alternative in the case where the mean of the five opinions was actually zero. First the program checked the response set of the unknown subject and recorded an answer of "slightly agree" or "slightly disagree" depending upon his tendency to favor responses on the agree or disagree side of the scale. If the response set was also zero, the program determined whether the majority of the group members thought that the unknown subject would answer on the agree side or the disagree side of the scale. An answer would be recorded that was in line with the majority opinion.

The members of the Austrian groups used a five point scale in giving their opinions, ranging from "Entschieden ja" through "Unentschieden" to "Entschieden nein". In this case the "undecided" category was allowed so that it was not necessary to use the response set and majority opinion routines to avoid the mid-point on the scale.

The second model used the mean of the individual opinions as the best estimate of the group decision, but this time members' opinions were given different weights according to their position in the communication network, their interaction weight, or their success with the task. Simulated group members also "learned" more about the unknown subject as the trials progressed by considering the extent to which he had been opinionated in his answers. The routines which simulate the effects of task success and learning, were primarily the work of Richardson.

The following aspects of member performance and subject behavior were incorporated into the program: absolute accuracy and relative accuracy of a member's predictions, and opinionatedness and response set of the subject's answers.

Absolute accuracy of members:

It was expected that the members of the group who were successful in predicting the unknown subject's responses on previous trials would have more influence in the discussion in subsequent trials. Also, the successful members would less readily change their opinions in order to conform to the existing group consensus, for they have achieved a certain amount of security in the group during previous trials and thus are more in a position to lead rather than to follow. In order to simulate this process, the "weights" of the successful members were increased for the subsequent trials.

It was also expected that the unsuccessful members would have less influence over the group and would more readily change their opinions to conform to the group consensus. This was simulated by decreasing the "weights" of the members of the group whose individual predictions differed greatly from the unknown subject's responses on the previous trials.

Relative accuracy of members:

The weights of the members were also changed according to the success of the members relative to the group. Thus, the weights were increased for those members who were more accurate than the group decision in predicting the unknown subject's responses. The weights were decreased for those members who were unsuccessful relative to the group's decision.

Uncertain vs. opinionated subjects:

As the groups were observed making actual decisions, it was apparent that the members often took into account whether or not the unknown subject tended to be uncertain or opinionated in his responses. If a group was trying to decide between a prediction of "strongly agree" or "agree", they would often choose the "strongly agree" response if the unknown subject's responses had tended to be in the "strongly" opinionated areas. In a similar way the group would have less tendency to choose the "strongly agree" or "strongly disagree" reponses if the unknown subject's response had tended to be in the "slightly" or uncertain areas. To simulate this learning process, the probability of the group decision falling into the "strongly" or "slightly" areas was increased or decreased according to the trend of the unknown subject's responses toward opinionatedness or uncertainty.

This learning process was also used to combat the tendency of "pooling" methods to arrive at decisions which cluster around the mean, or uncertain area. If an individual has learned in previous trials that the unknown subject has tended to be rather opinionated, he is more likely to have an individual prediction of either "strongly agree" or "strongly disagree" on succeeding trials.

For a given item on the value-orientation test this member may not be able to decide, before the group decision takes place, whether the true response will be in the agree or disagree categories. He may think that the true response is probably either "strongly agree" or "strongly disagree". Since he must make a decision he decides to guess "strongly disagree". However, if the others in the group feel that the true response is at the "agree" end of the scale, he will be more likely to go along with a group decision of "strongly agree" rather than "slightly agree", even though a group decision of "strongly agree" is furthest away from his original prediction.

Response set of subject:

As in the first model, the response set of the unknown subject was used for the American groups whenever the pooling of individual opinions indicated that the group had "no opinion" for a given trial. The limits of the "no opinion" area were determined empirically in model two and were only one-fifth as wide as those used in model one. In other respects the use of the response set was the same as before. When the response set was neither positive nor negative, the simulated group decision was taken to be in the direction of the majority opinion for that trial.

In adapting this program to the five point scale used by the Austrian groups the routines used to deal with group guesses of "no opinion" were again eliminated since the "undecided" category was allowed. An additional modification in the simulation which had not been tried before with any of the groups was to change the limits used to determine which of the possible answers the mean was assumed to represent. This was done to increase the probability of giving a simulated group prediction in the "agree" or "disagree" categories for the American groups (i.e., answers of +2 or -2) and increasing the probability of simulating the most extreem categories for the Austria groups.

Using the second model the accuracy of the simulation was compared fpr all groups with and without "learning" for each of the following conditions: all group members' opinions given the same weight, members given weights according to their total interaction during the session, and members weighted according to total interaction using the wider limits for predicting extreem answers.[3)]

The third model, which was developed in Vienna, assumes that the majority opinion or the median is the best predictor of group decisions. In the case of a five man group, a majority of three also contains the median.

The decision to use the median grew out of a discussion between Hare and Scheiblechner as they attempted to find the reasons why the previous simulations failed to predict more of the group decisions. Hare proposed using the majority opinion and Scheiblechner the median. The program written by Hare is actually the one used in the third model. It begins finding a majority of three when there is one for each trial. However when there is no majority the median is used. In a few cases, in the American groups, where the lack of a recorded opinion for some individuals would place the median in the "no opinions" category, a majority of two is used. Using this model in a five man group it is never necessary to use any of the "learning" features of model two, nor do the "weights" assigned the subjects make any difference.

Sources of data

The groups used to test the simulations come from three sources: Havard, Haverford and Villanova, and Vienna. The twenty five-man groups at Harvard were observed as part of an experiment conducted by Churchill (1961). The five Haverford and Villanova groups were observed by Hare and Richardson, and the six Austrian groups were observed at the Institute for Advanced Study in Vienna by Scheiblechner.[4)]

In the majority of the groups the subjects were male undergraduates.

The data for the American groups consist of the individual predictions of the unknown subject's responses to the Bales-Couch Value Profile (1960), the group decisions, and the actual responses of the unknown subjects. Two unknown subjects were used, one an actual "unpredictable" graduate student for the Havard groups, and the other a fabricated "predictable man" for the Haverford and Villanova groups.

The data for the Austrian groups are similar with the substitution of a values test, in German, developed by Reichardt, Prof. of Sociology, Head of the Institute of Sociology, University of Vienna. Here a fabricated unknown subject was used who turned out to be relatively unpredictable.

In both countries the individuals' total interaction rates for the entire session and seating positions were recorded. Each five man group completed roughly ten trials. The Havard experiment was conducted in a small group laboratory using an observation mirror and the other groups were observed in classrooms.

## The computer programs

The program for model one was written in machine language for the IBM 650. For model two the program was first written in SPS for an IBM 1620 and then translated to FORTRAN. Model three programs and adaptations of earlier programs were done in FORTRAN and rund on an IBM 1620 Mark II. None of the programs requires a very elaborate computer.

As described above, the basic method of simulating a group's decision in Hare's first (p.1) model was to take the unweighted mean of the individual opinions. For model two, various weights which reflected interaction rate or seating position were used as the "starting weights", i.e., the weights used for the first trial before any learning process had taken place. For interaction rate we predicted that those who talked the most would have the most influence and for seating position, those in seats 1 and 5 at the ends of the table, and 3 in the center of the side facing the experimenter.

Also in model two routines were added to test the hypotheses concerning the relative influence of successful members and the learning process. The weights for the individuals were increased (or decreased) by a certain percentage, depending upon the individual's success (or failure), both absolutely and relatively, on the preceeding trial. With the addition of new information about the unknown subject's actual response for each trial the proba'ilities for opionionated or uncertain opinions being given as the group predictions were changed. For model three the mode, or median if there was no mode, was selected without the "learning" routines.

## Results

The results for the various simulations with the 179 trials of the Harvard groups are given in Table 1, for the 68 trials of the combined Haverford and Villanova groups in Table 2, and for the total of 247 trials with the American sample in Table 3. The data for the 60 trials in the Austrian sample which used a different value test and a narrower range of permitted responses are given in Table 4.

Table 1

HARVARD GROUPS
SIMULATION OF GROUP DECISIONS

(20 GROUPS, 179 TRIALS)

| SIMULATION TYPE | | DIFFERENCE BETWEEN SIMULATION AND ACTION GROUP DECISION | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | TOTAL |
| MEAN: Unit Weights | No Learning | 105 | 57 | 15 | 2 | 179 |
| | Learning | 114 | 45 | 18 | 2 | |
| MEAN: Seating Weights | No Learning | 104 | 52 | 19 | 4 | |
| | Learning | 118 | 39 | 17 | 5 | |
| MEAN: Interaction Weights | No Learning | 113 | 49 | 15 | 2 | |
| | Learning | 119 | 42 | 14 | 4 | |
| MEAN: Wide Cutting Points (Int.WTS) | No Learning | 119 | 47 | 10 | 3 | |
| | Learning | 125+) | 35 | 13 | 6 | |
| MODE or MEDIAN (WTS. NOT USED) | | 133++) | 35 | 7 | 4 | |

(January 30, 1968)

+) Significant increase over unit weights - no learning $p < 05$ Using Sign Test

++) $p < 01$

Table 2

HAVERFORD AND VILLANOVA GROUPS
(PREDICTABLE MAN)

(5 GROUPS , 68 TRIALS)

| SIMULATION TYPE | | DIFFERENCE BETWEEN SIMULATION AND ACTUAL GROUP PENSION | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | TOTAL |
| MEAN: Unit Weights | No Learning | 34 | 28 | 6 | | |
| | Learning | 36 | 25 | 6 | 1 | 68 |
| MEAN: Seating Weights | No Learning | 34 | 25 | 8 | 1 | |
| | Learning | 32 | 25 | 9 | 2 | |
| MEAN: Interaction WTS | No Learning | 34 | 28 | 6 | 0 | |
| | Learning | 35 | 26 | 4 | 3 | |
| MEAN: Wide Cutting Points (Int. WTS) | No Learning | 40 | 23 | 4 | 1 | |
| | Learning | 40 | 21 | 2 | 5 | |
| MODE OR MEDIAN (WTS not used) | | 48 +) | 13 | 6 | 1 | |

+) Significant increase over unit weights -
no learning using sign test $p < 05$

Table 3

TOTAL HARVARD, HAVERFORD AND VILLANOVA GROUPS
(25 GROUPS, 247 TRIALS)

| SIMULATION TYPE | | DIFFERENCE BETWEEN SIMULATION AND ACTUAL GROUP DECISION | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | TOTAL |
| MEAN: Unit Weights | No Learning | 139 | 85 | 21 | 2 | 247 |
| | Learning | 150 | 70 | 24 | 3 | |
| MEAN: Seating Weights | No Learning | 138 | 77 | 27 | 4 | |
| | Learning | 150 | 64 | 26 | 7 | |
| MEAN: Interaction Weights | No Learning | 147 | 77 | 21 | 2 | |
| | Learning | 154 | 68 | 18 | 7 | |
| MEAN: Wide Cutting Points (Int. Weights) | No Learning | 159 +) | 70 | 14 | 4 | |
| | Learning | 165 +) | 56 | 15 | 11 | |
| MODE OR MEDIAN (Wts. not used) | | 181 ++) | 48 | 12 | 6 | |

+) Significant increase over unit weights - no learning using sign test $p \leq 05$

++) $p \leq 001$

Table 4

VIENNA GROUPS

SIMULATION OF GROUP DECISIONS

(6 Groups x 10 Trials = 60 Trials)

| SIMULATION TYPE | | DIFFERENCE BETWEEN SIMULATION AND ACTUAL GROUP DECISION | | | |
|---|---|---|---|---|---|
| | | 0 | 1 | 2 | TOTAL |
| MEAN: Unit Weights | No Learning | 44 | 16 | | 60 |
| | Learning | 41 | 13 | | |
| MEAN: Seating Weights | No Learning | 44 | 16 | | |
| | Learning | 42 | 17 | 1 | |
| MEAN: Interaction Weights | No Learning | 47 | 13 | | |
| | Learning | 48 | 11 | 1 | |
| MEAN: Wide Cutting Points (Interaction weights) | No Learning | 47 | 13 | | |
| | Learning | 47 | 12 | 1 | |
| MODE OR MEDIAN (Weights not used) | | 56 | 4 | | |

With the first model, taking the unweighted mean, 105 trials out of 179 are correctly simulated for the Havard groups (Table 1), 34 out of 68 for the Haverford-Villanova groups (Table 2), and 44 out of 60 for the Vienna groups. Thus more than half of the decisions are accurately simulated by the simplest model.[5)]

When the various starting weights are used as part of the second model, there is no improvement using the seating weights and only slight improvement using interaction weights in the Harvard and Vienna groups. However this apparent increase is not statistically significant.[6)]

For both American and Austrian groups there is some increase in the accuracy of the simulation when the learning routines are added to the unit weights and the interaction. However the increase over the simplest model of unit weights and no learning is not significant.

The introduction of the wider cutting points for the more extreem answers improves the accuracy of the prediction over model one in all cases although learning only adds to the accuracy for the American groups. The improvement for the wider cuttings points, with and without learning, for the combined American groups is significant at the .05 level.

Finally the use of model three, the median, make a significant improvement for both sets of American groups at least at the .05 level, and for the combined American groups at the .01 level. Although there is an increase of accuracy of the simulation in five of the six Austrian groups, the probability that this could happen by chance is more than .05 for a sample of this size.

For ten of the twelve cases, the shift from a "no learning" to a "learning" model, which is usually associated with an increase in the accuracy of the prediction of the actual answer, also results in a greater number of gross errors (See Tables 1,2 and 4). The most extreem case of this is found in Table 3 when the wide cutting points are used. With no learning there is only one simulation which differes by three points from the actual group decision, whereas with learning the number of errors of this size rises to five. Thus absolute accuracy is purchased at the price of greater relative error when the simulation is incorrect. This is also true for the American groups when the simulation under model three is compared with model one. What is happening here is that the learning routines, the wider cutting points, or taking the median rather than the mean, all have the effect of making a somewhat more extreem prediction, which if it is right, reduces the descrepancy with the actual group guess to zero, but if it is wrong, increases the number of large errors.

Summary

In a test of three computer models to simulate group decisions, data were used from 31 American and Austrian groups on a total of 307 trials. The task for each group was to predict a series of answers of an unknown subject on a value-orientation questionnaire after being given a sample of his typical responses.

The first model which used the mean of the individual opinions as a simulation of the group judgement simulated exactly over half of the trials. The simulation was improved in model two, which also used the mean, when individual opinions were weighted according to their total participation in the discussion and "learning" was added. This was especially true when the cutting points were changed so that the mean would represent more extreem opinions.

The best simulation occured with model three which used the median of the individual opinions as the best simulation of the group decisions. Using this model over 75 per cent of the trials were accurately simulated.

5) In earlier versions of models one and two the published results differed slightly as a result of minor changes in programming as we used new programming languages and machines. Although some of the results of the simulations appeared more orderly in previous tables, they were not significantly different from the present figures.

6) Test of significance were made by comparing the increase in absolute prediction group by group under each condition using a sign test (Siegel 1956, Table D). If an increase in simulation was observed it was scored as plus, if there was no increase, scored as zero, and a decrease scored as minus. The probability of the number of minuses given the total number of groups in which a change was observed was noted in Table D. As a further test the correlations between the simulated and actual group guesses were computed for each group and the average correlations compared, after using the z' transformation. The results were similar (See Appendix)

References and Sources

1. Bales, R.F., The equilibrium problem in small groups. In T. Parsons, R.F.Bales, & E.A.Shils, Working papers in the theory of action. New York: Free Press of Glencoe, 1953. pp. 111-161.

2. Bales, R.F. Interaction process analysis. In D.L.Sills (Ed.), International Encyclopedia of the Social Sciences, New York: Macmillan, Free Press, and Collier's Encyclopedia, 1968

3. Bales, R.F. & Couch, A.S., The value profile: A factor analytic study of value statements. Harvard University: Laboratory of Social Relations, (ditto), 1960.

4. Bales, R.F., Couch, A.S., & Stone, P.J. The interaction simulator.In proceedings of a symposium on digital computers and their applications, Annals of the Computation Laboratory, (Harvard University), Vol. 31,305-314, c.1961:

5. Churchill, L.C., Jr., Aggression in a small group setting. Unpublished doctoral dissertation, Harvard, 1960.

6. Couch, A.S. Psychological determinato of interpersonal behavior, Unpublished Doctoral Dissertation, Department of Social Relations, Harvard University, 1960.
   Stone, P.J. et al General Inquirer ........... 1966

7. Hare, A.P. Computer simulation of interaction in small groups. Behavioral Science, 1961, 6, pp. 261-265

8. Hare, A.P., Handbook of small group research. New York: Free Press of Glencoe, 1962.

9. Hare, A.p., & Bales, R.F., Seating position and small group interaction. Sociometry, 1963, 26, pp. 480-486.

10. Hare, A.p. Richardson, R. Computer simulation of small group decisions. Paper presented at meetings of American Sociological Association, Miami, Florida, August 1966.

## References

11. Kaplan, A., Skogstad, A.L., & Girshick, M.A., The prediction of social and technological events. Publ. Opin. Quart. 1950, 14, pp. 93-110.

12. Siegel, S. Nonparametric statistics, Mac Grew Hill, 1956.

13. Stone, P. & Kamiya, J. Judgement of consensus during group discussion. J. abnorm. soc. Psychol., 1957, 55, pp. 171-175

14. Stone Ph. J., Dunphy D.C., Smith M.S., Ogilvie D.M. The General Inquirer: A Computer Approach to Content Analysis, Cambridge, Mass. MIT, 1966

## Appendix

Test of simulations via correlation coefficients

For each group the product-moment correlation between simulated and actual group decisions was computed. Then the r's were transformed according to the z-transformation and the homogeneity of the coefficients tested. After that, mean correlations were calculated. The following table shows the mean correlations for the various samples of groups under the four simulations. The coefficients derived from heterogeneous samples are put in brackets (they are not strictly statistically legitimate)

Mean correlations between simulated and actual group-decision

| | Simulations | | | |
|---|---|---|---|---|
| Origin of groups | Unit Weights NL | Interaction Weights L | Wider Limits L | Median |
| 20 Harvard groups | 0,949 | 0,949 | 0,937 | 0,967 |
| 3 Villanova gps. | (0,94) | 0,88 | (0,903) | (0,954) |
| 2 Haverford gps. | 0,932 | 0,90 | 0,908 | 0,933 |
| 6 IHS grps. | (0,92) | (0,928) | 0,902 | (0,978) |
| Total | (0,941) | (0,933) | 0,924 | 0,965 |

For the twenty Harvard groups the difference between all other simulations and the median is significant at the .05-level ($\chi^2$ = 4,014, df=1). For the total groups the difference between the "Wide limits-L" and the Median is significant at far beyond the .001-level ($\chi^2$ = 23,4 df = 1).